

## Math 117 – Summarizing section 4.1 – Writing hypotheses

---

- 1) Write the hypotheses; give the notation for a sample statistic we might record to test the hypotheses; will the test be left, right or two tailed?  
What will the center of the randomization distribution be?

4.5 - Testing to see if there is evidence that the mean of group A is not the same as the mean of group B.

4.6 - Testing to see if there is evidence that a proportion is greater than 0.3.
















4.7 - Testing to see if there is evidence that a mean is less than 50.

4.9 - Testing to see if there is evidence that the proportion of people who smoke is greater for males than for females.

## Section 4.2 MEASURING EVIDENCE WITH P-VALUES

### SECTION LEARNING GOALS

*You should now have the understanding and skills to:*

- Recognize that a randomization distribution shows what is likely to happen by random chance if the null hypothesis is true  
 [Video Tutorial: Learning Goal 4.2.1](#)  
- Use technology to create a randomization distribution  
 [Video Tutorial: Learning Goal 4.2.2](#)  
- Interpret a p-value as the proportion of samples that would give a statistic as extreme as the observed sample, if the null hypothesis is true  
 [Video Tutorial: Learning Goal 4.2.3](#)  
- Distinguish between one-tailed and two-tailed tests in finding p-values  
 [Video Tutorial: Learning Goal 4.2.4](#)  
- Find a p-value from a randomization distribution  
 [Video Tutorial: Learning Goal 4.2.5](#)  

### Randomization Distribution

Simulate many samples assuming the null hypothesis is true. Collect the values of a sample statistic for each simulated sample to create a **randomization distribution**.

The randomization distribution will be centered at the value indicated by the null hypothesis and shows what values of the sample statistic are likely to occur by random chance, if the null hypothesis is true.

## Section 4.2 MEASURING EVIDENCE WITH P-VALUES

### 2) Does Light at Night Affect Weight Gain?

Numerous studies have shown that exposure to light at night is harmful to human health. A recent study<sup>2</sup> examines the effect of light at night on body mass gain in mice. Eighteen mice were randomly assigned to one of two groups: the *Dark* group had a normal light/dark cycle with darkness at night and the *Light* group had a dim light on at night, equivalent to having a television set on in a room.<sup>3</sup> The body mass gain (*BMGain*), in grams, was recorded after three weeks, and the results are given in **LightatNight** and in Table 4.1. Do the data in this study provide convincing evidence that having a light on at night increases weight gain in mice? **NOTE THAT  $\bar{X}\text{-BAR-L} = 6.732$  and  $\bar{X}\text{-BAR-D} = 4.114$  gr (data is given in the book, section 4.1, page 258). We wonder: is  $\bar{x}\text{-bar-L}$  higher just by chance or is it an indication that light increases the average BMI?**

- a) Write the hypotheses; what tail test is this? Sketch randomization distribution; label center. Find the sample statistic and label the possible positions on the graph.

Go to STATKEY and generate the randomization distribution for this example. Click on the appropriate tail-test and locate the sample statistic in your randomization. According to your randomization, what is the “likelihood” of observing such a sample statistic or a more extreme one? (this is the p-value).

**The P-value measures how extreme sample results would be, if the null hypothesis were true.**

The **p-value** is the proportion of samples, when the null hypothesis is true, that would give a statistic as extreme as (or more extreme than) the observed sample.

**Read the book – explanation about obtaining P-values from Randomization Distributions – take notes**

## Section 4.2 – Book Problems

### 3) What tail test? Sketch and label.

#### SKILL BUILDER 2

In Exercises 4.45 to 4.49, the null and alternative hypotheses for a test are given as well as some information about the actual sample(s) and the statistic that is computed for each randomization sample. Indicate where the randomization distribution will be centered. In addition, indicate whether the test is a left-tail test, a right-tail test, or a two-tailed test.

4.45 Hypotheses:  $H_0 : p = 0.5$  vs  $H_a : p < 0.5$

Sample:  $\hat{p} = 0.4$ ,  $n = 30$

Randomization statistic =  $\hat{p}$

ANSWER ⊕

WORKED SOLUTION ⊕

4.46 Hypotheses:  $H_0 : \mu = 10$  vs  $H_a : \mu > 10$

Sample:  $\bar{x} = 12$ ,  $s = 3.8$ ,  $n = 40$

Randomization statistic =  $\bar{x}$

---

### 4) What is the randomization statistic? Sketch and label.

4.48 Hypotheses:  $H_0 : \mu_1 = \mu_2$  vs  $H_a : \mu_1 \neq \mu_2$

Sample:  $\bar{x}_1 = 2.7$  and  $\bar{x}_2 = 2.1$

Randomization statistic =  $\bar{x}_1 - \bar{x}_2$

4.49 Hypotheses:  $H_0 : p_1 = p_2$  vs  $H_a : p_1 > p_2$

Sample:  $\hat{p}_1 = 0.3$ ,  $n_1 = 20$  and  $\hat{p}_2 = 0.167$ ,  $n_2 = 12$

Randomization statistic =  $\hat{p}_1 - \hat{p}_2$

## Section 4.2 – Book Problems

### 5) How likely are the sample statistics?

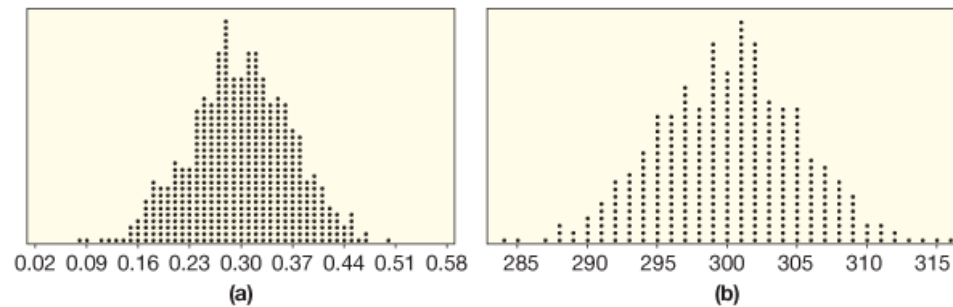
In Exercises 4.50 and 4.51, a randomization distribution is given for a hypothesis test, and shows what values of the sample statistic are likely to occur if the null hypothesis is true. Several possible values are given for a sample statistic. In each case, indicate whether seeing a sample statistic as extreme as the value given is (i) reasonably likely to occur when the null hypothesis is true, (ii) unusual but might occur occasionally when the null hypothesis is true, or (iii) extremely unlikely to ever occur when the null hypothesis is true.

**4.50** Figure 4.13(a) shows a randomization distribution for a hypothesis test with  $H_0 : p = 0.30$ . Answer the question for these possible sample proportions:

(a)  $\hat{p} = 0.1$

(b)  $\hat{p} = 0.35$

(c)  $\hat{p} = 0.6$



**Figure 4.13** Randomization distributions for Skill Builder 3

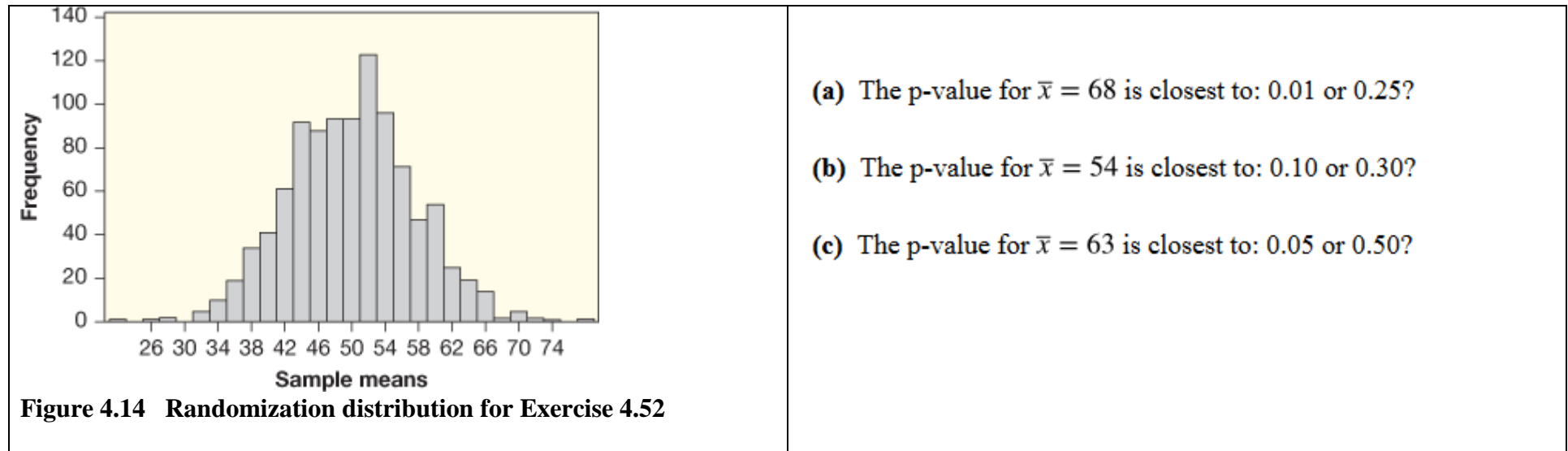
### 6) How likely are the sample statistics?

**Exercise 4.51 of the book deals with graph (b)** shown above. (complete it at home)

- Write the null hypothesis.
- Use the graph to give examples of sample statistics that:
  - reasonably likely to occur when the null hypothesis is true,
  - unusual but might occur occasionally when the null hypothesis is true, or
  - extremely unlikely to ever occur when the null hypothesis is true.

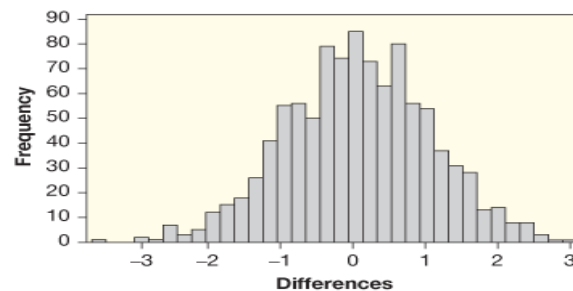
## Section 4.2 – Book Problems - Which p-value most closely matches the observed statistic?

- 7) 4.52 - Figure 4.14 shows a randomization distribution based on 1000 simulated samples for testing  $H_0 : \mu = 50$  vs  $H_a : \mu > 50$ . Use the distribution to decide which p-value most closely matches the observed statistic?



## 8) Which p-value most closely matches the observed statistic?

- 4.54 Figure 4.16 shows a randomization distribution for testing  $H_0 : \mu_1 = \mu_2$  versus  $H_a : \mu_1 \neq \mu_2$ . The statistic used for each sample is  $D = \bar{x}_1 - \bar{x}_2$ . In each case, use the distribution to decide which value is closer to the p-value for the observed difference in sample means.



**Figure 4.16** Randomization distribution for Exercises 4.54 to 4.58

- (a) The p-value for  $D = \bar{x}_1 - \bar{x}_2 = -2.9$  is closest to: 0.01 or 0.25?
- (b) The p-value for  $D = \bar{x}_1 - \bar{x}_2 = 1.2$  is closest to: 0.30 or 0.60?

## Section 4.2 – Book Problems

### 9) Which sample statistic provides the strongest evidence against $H_0$ ?

#### SKILL BUILDER 5

Exercises 4.55 to 4.58 also refer to Figure 4.16, which shows a randomization distribution for hypotheses  $H_0 : \mu_1 = \mu_2$  vs  $H_a : \mu_1 \neq \mu_2$ . The statistic used for each sample is  $D = \bar{x}_1 - \bar{x}_2$ . Answer parts (a) and (b) using the two possible sample results given in each exercise.

- (a) For each  $D$ -value, sketch a smooth curve to roughly approximate the distribution in Figure 4.16, mark the  $D$ -value on the horizontal axis, and shade in the proportion of area corresponding to the p-value.
- (b) Which sample provides the strongest evidence against  $H_0$ ? Why?

4.55  $D = 2.8$  or  $D = 1.3$

4.56  $D = 0.7$  or  $D = -1.3$

4.57  $\bar{x}_1 = 17.3, \bar{x}_2 = 18.7$  or  $\bar{x}_1 = 19.0, \bar{x}_2 = 15.4$

4.58  $\bar{x}_1 = 95.7, \bar{x}_2 = 93.5$  or  $\bar{x}_1 = 94.1, \bar{x}_2 = 96.3$

---

### 10) Using STATKEY to generate a randomization distribution

#### SKILL BUILDER 6

Exercises 4.59 to 4.64 give null and alternative hypotheses for a population proportion, as well as sample results. Use *StatKey* or other technology to generate a randomization distribution and calculate a p-value. *StatKey* tip: Use “Test for a Single Proportion” and then “Edit Data” to enter the sample information.

4.59 Hypotheses:  $H_0 : p = 0.5$  vs  $H_a : p > 0.5$

Sample data:  $\hat{p} = 30/50 = 0.60$  with  $n = 50$

4.60 Hypotheses:  $H_0 : p = 0.5$  vs  $H_a : p < 0.5$

Sample data:  $\hat{p} = 38/100 = 0.38$  with  $n = 100$

4.61 Hypotheses:  $H_0 : p = 0.7$  vs  $H_a : p < 0.7$



















Sample data:  $\hat{p} = 125/200 = 0.625$  with  $n = 200$

---

## Section 4.3 - DETERMINING STATISTICAL SIGNIFICANCE

### SECTION LEARNING GOALS

*You should now have the understanding and skills to:*

- Recognize that smaller p-values give stronger evidence in support of the alternative hypothesis  
 [Video Tutorial: Learning Goal 4.3.1](#)  
- Demonstrate an understanding of the concept of statistical significance  
 [Video Tutorial: Learning Goal 4.3.2](#)  
- Make a formal decision in a hypothesis test by comparing a p-value to a given significance level  
 [Video Tutorial: Learning Goal 4.3.3](#)  
- State the conclusion to a hypothesis test in context  
 [Video Tutorial: Learning Goal 4.3.4](#)  
- Make a less formal decision that reflects the strength of evidence in a p-value  
 [Video Tutorial: Learning Goal 4.3.5](#)  
- Conduct a hypothesis test for a variety of situations  
 [Video Tutorial: Learning Goal 4.3.6](#)  



## Section 4.3 - DETERMINING STATISTICAL SIGNIFICANCE

### Interpreting a P-value as Strength of Evidence

The smaller the p-value, the stronger the statistical evidence is against the null hypothesis and in support of the alternative hypothesis.

### Statistical Significance

If the p-value is small enough, then results as extreme as the observed sample statistic are unlikely to occur by random chance alone (assuming the null hypothesis is true), and we say the sample results are **statistically significant**.

If our sample is statistically significant, we have convincing evidence against  $H_0$  and in favor of  $H_a$ .

### 1) Which provides the strongest evidence against $H_0$ ?

#### SKILL BUILDER 1

In Exercises 4.80 to 4.83, two p-values are given. Which one provides the strongest evidence against  $H_0$ ?

**4.80** p-value = 0.90 or p-value = 0.08

**4.81** p-value = 0.04 or p-value = 0.62

ANSWER ⊕

WORKED SOLUTION ⊕

**4.82** p-value = 0.007 or p-value = 0.13

**4.83** p-value = 0.02 or p-value = 0.0008

## Section 4.3 - DETERMINING STATISTICAL SIGNIFICANCE

### Significance Level

The **significance level**,  $\alpha$ , for a test of hypotheses is a boundary below which we conclude that a p-value shows statistically significant evidence against the null hypothesis.

Common significance levels are  $\alpha = 0.05$ ,  $\alpha = 0.01$ , or  $\alpha = 0.10$ . If a significance level is not specified, we use  $\alpha = 0.05$ .

### Formal Statistical Decisions

Given a significance level  $\alpha$  and the p-value from a sample:

If the p-value  $< \alpha$ : *Reject  $H_0$* . This means the results are significant and we have convincing evidence that  $H_a$  is true.

If the p-value  $\geq \alpha$ : *Do Not Reject  $H_0$* . This means the results are not significant and we do not have convincing evidence that  $H_a$  is true.

## 2) Reject or do not reject the Null?

### SKILL BUILDER 2

Exercises 4.84 to 4.87 give a p-value. State the conclusion of the test based on this p-value in terms of “Reject  $H_0$ ” or “Do not reject  $H_0$ ”, if we use a 5% significance level.

**4.84** p-value = 0.0007

**4.85** p-value = 0.0320

ANSWER 

WORKED SOLUTION 

**4.86** p-value = 0.2531

**4.87** p-value = 0.1145

**3) Are results significant?**

**SKILL BUILDER 3**

In Exercises 4.88 to 4.91, using the p-value given, are the results significant at a 10% level? At a 5% level? At a 1% level?

**4.88** p-value = 0.0320

**4.89** p-value = 0.2800

**4.90** p-value = 0.008

**4.91** p-value = 0.0621

---

**4) Significance and conclusions**

**SKILL BUILDER 4**

In Exercises [4.92](#) and [4.93](#), match the four p-values with the appropriate conclusion:

- (a) The evidence against the null hypothesis is significant, but only at the 10% level.
- (b) The evidence against the null and in favor of the alternative is very strong.
- (c) There is not enough evidence to reject the null hypothesis, even at the 10% level.
- (d) The result is significant at a 5% level but not at a 1% level.

**4.92 I** 0.0875

**II** 0.5457

**III** 0.0217

**IV** 0.00003

### Section 4.3 - DETERMINING STATISTICAL SIGNIFICANCE

#### **Hypothesis Tests**

A formal hypothesis test includes the following components:

- •State the null and alternative hypotheses (defining parameters when necessary)
- •Determine the value of the observed sample statistic
- •Find the p-value
- •Make a generic decision about  $H_0$  : Reject  $H_0$  or do not reject  $H_0$
- •Write a sentence explaining the conclusion of the test in context, indicating whether or not we have convincing evidence for  $H_a$  and referring back to the question of interest.

## Section 4.3 – What does it mean Statistically Significant?

5) **Statistical Significance** - What does it mean **statistically significant**? Circle all the correct choices below:

- i. We have convincing evidence in favor of  $H_0$  and against  $H_a$
- ii. We have convincing evidence against  $H_0$  and in favor of  $H_a$
- iii. When results as extreme as the ones observed are unlikely to happen by random chance alone (assuming  $H_0$  is true)
- iv. When results as extreme as the ones observed are likely to happen by random chance alone (assuming  $H_0$  is true)

---

6) In each pair of statements, select the correct one.

- a) What does a very small p-value mean?
  - i. The sample is very likely to occur by random chance alone when the null is true
  - ii. The sample is very unlikely to occur by random chance alone when the null is true
  
- i. Test results are not statistically significant
  - ii. Test results are statistically significant
  
- i. The test statistic is far into the tail in the direction of the alternative
  - ii. The test statistic is close to the center of the distribution of  $H_0$
  
- i. There is strong evidence against the null hypothesis,  $H_0$ , in favor of the alternative,  $H_a$
  - ii. There is strong evidence in favor of the null hypothesis,  $H_0$ , and against the alternative,  $H_a$
  
- b) If the evidence is strong enough against the null hypothesis, which of the following is correct?
  - i. We can *reject the null hypothesis* in favor of the alternative.
  - ii. We can *reject the alternative hypothesis* in favor of the null.
  
- c) If the data are reasonably likely to occur when the null hypothesis is true,
  - i. We *do not reject the null hypothesis*.
  - ii. We *reject the null hypothesis*.

**IMPORTANT NOTE:** If the data are not significant, we do not reject  $H_0$ ; however, we do not say that we “accept  $H_0$ .”

### 4.3 – p-values - Reject $H_0$ or Do not Reject $H_0$ ? Writing conclusions in context.

#### 7) Example: *Red Wine and Weight Loss*

Resveratrol, an ingredient in red wine and grapes, has been shown to promote weight loss in animals. In one study, a sample of lemurs had various measurements taken before and after receiving resveratrol supplements for 4 weeks.

For each situation described below, write the hypotheses. Sketch, label and shade. Based on the given p-value, indicate whether you Reject or do not reject  $H_0$ . Write the conclusion in context. Use a 5% significance level.

a) In the test to see if the mean resting metabolic rate is higher after treatment, the p-value is 0.013.

b) In the test to see if the mean body mass is lower after treatment, the p-value is 0.007.

c) In the test to see if mean locomotor activity changes after treatment, the p-value is 0.980.

d) In the test to see if mean food intake changes after treatment, the p-value is 0.035.

e) Which of the results given in (a) – (d) above are significant at a 1% level?

#### 4.3 – p-values. Reject $H_0$ or Do not Reject $H_0$ ? Writing conclusions in context.

##### **8) Example 2: Multiple Sclerosis and Sunlight**

It is believed that sunlight offers some protection against multiple sclerosis, but the reason is unknown. Is it the vitamin D, the UV light, or something else? In an experiment, mice were injected with a substance to give them MS and were randomly assigned to either a control group (with no treatment), a group that received vitamin D supplements, or a group that got exposed regularly to UV light. The scientists found that mice exposed to UV light were significantly less likely to get MS than the control mice, but that vitamin D did not seem to reduce the likelihood of getting MS compared to the control group. For these two tests, one of the p-values was 0.470 and one was 0.002. Which p-value goes with which test? Also, for each test, indicate whether we “Reject  $H_0$ ” or “Do not reject  $H_0$ ” and write the conclusions in context.

## 4.2, 4.3 – p-values. Reject $H_0$ or Do not Reject $H_0$ ? Writing conclusions in context.

### 9) Support for the Death Penalty

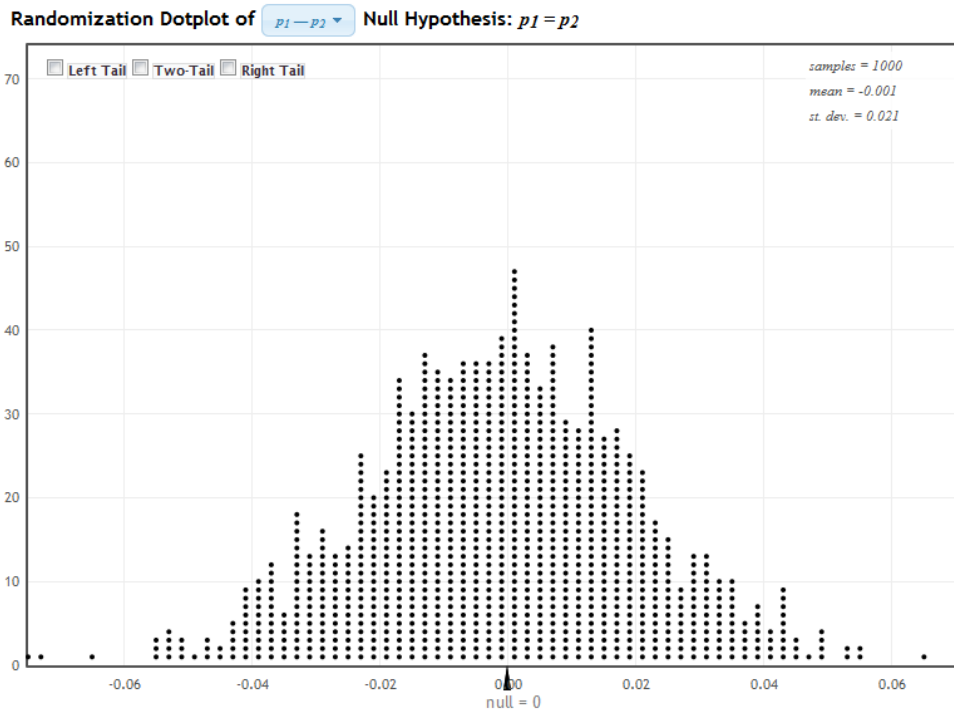
In 1980 and again in 2010, a Gallup poll asked a random sample of 1000 US citizens “Are you in favor of the death penalty for a person convicted of murder?” In 1980, the proportion saying yes was 0.66. In 2010, it was 0.64. Does this data provide evidence that the proportion of US citizens favoring the death penalty was higher in 1980 than it was in 2010? Use  $p_1$  for the proportion in 1980 and  $p_2$  for the proportion in 2010.

#### First let's reflect about the problem

- a) Who/what are the cases?   b) One or two groups?   c) Variable?   d) Categorical or quantitative?  
e) Objective of the study (circle in the statement of the problem)   f) What is this problem about and what notation will be used?  
(1) One single mean?   (2) One single proportion?  
(3) Difference of means?   (4) Difference of proportions?

a) State the null and alternative hypotheses:

b) Give the value of the sample statistic    $\hat{p}_1 - \hat{p}_2$    and locate in the graph below.



c) A randomization distribution assuming the null hypothesis is true is shown. Which of the following is closest to the p-value?

0.001,   0.05,   0.20,   0.5

d) Circle the correct statement:

(1) Results are statistically significant

(2) Results are not statistically significant

e) TRUE or FALSE? We have strong evidence against the null hypothesis and in favor of the alternative.

f) Write the conclusion in context.

g) What is the p-value for a sample statistic of 0.04? Will the conclusion be the same as in part (f)?



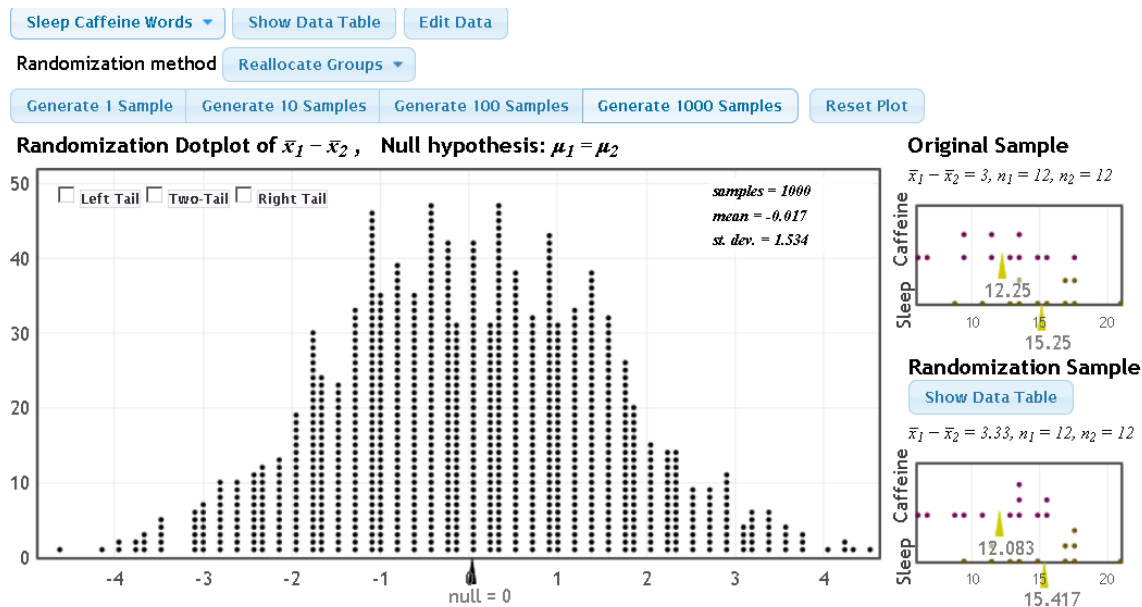
## 4.2, 4.3 – p-values. Reject $H_0$ or Do not Reject $H_0$ ? Writing conclusions in context.

### 10) Sleep vs Caffeine for Memory

In an experiment, students were given words to memorize, then were randomly assigned to either take a 90-minute nap or take a caffeine pill. A couple hours later, they were tested on their recall ability. We wish to test to see if the sample provides evidence that there is a difference in the mean number of words people can recall depending on whether they take a nap or have some caffeine.

#### First let's reflect about the problem

- a) Who/what are the cases?   b) One or two groups?   c) Variable?   d) Categorical or quantitative?  
e) Objective of the study (circle in the statement of the problem)   f) What is this problem about and what notation will be used?  
(1) One single mean?   (2) One single proportion?  
(3) Difference of means?   (4) Difference of proportions?



- a) Write the hypotheses:
- b) The observed sample statistic is  $\bar{x}_s - \bar{x}_c = 3.0$ . Use the randomization distribution to state the p-value.
- c) Under the rule: "If  $p < 0.05$  we have statistically significant results". Are the results of this test statistically significant? Write the conclusion in context.
- d) My perception is that "a nap helps with word recall". Write the hypotheses in this case, state the p-value and write the conclusion in context.